Studies in Second Language Acquisition, 2011, **33**, 529–561.
doi:10.1017/S0272263111000301

SECOND LANGUAGE ASSESSMENT AND MORPHOSYNTACTIC DEVELOPMENT

Patti Spinner
Michigan State University

The purpose of this study is to begin work toward a grammatical assessment measure that could bridge the gap between theoretical work on grammatical development, on the one hand, and tools such as the Michigan Test (which uses multiple-choice questions on vocabulary and grammar) or the American Council on the Teaching of Foreign Languages scale (which uses holistic descriptions of grammar use), on the other hand. Such a measure would need to be practical to administer with large groups. Two proposals of grammatical development (processability theory, Pienemann, 1998, 2005; and organic grammar, Vainikka & Young-Scholten, 2006) were applied to short samples of spontaneous production data from 48 adult second-language learners of English from mixed first-language backgrounds. The rapid profile scale successfully accounted for the learners' development but is of somewhat limited use with short samples of data. The organic grammar placement scale may need to be further refined, but it includes important indicators of grammatical development. A preliminary proposal for using a combined measure with a rubric is presented.

It has long been recognized that there is a fundamental connection between SLA research and second language (L2) assessment (e.g., Bachmann, 1990, 1998; Bachman & Cohen, 1998; Brindley, 1998; Chapelle, 1998; Cohen, 1998;

Part of this article was presented at EuroSLA 2007. I would like to thank Alan Juffs, Shawn Loewen, Pascual Masullo, Charlene Polio, Yasuhiro Shirai, Dick Tucker, Paula Winke, and several anonymous *SSLA* reviewers for insightful comments on various aspects of this article. All errors of course remain my own.

Address correspondence to: Patti Spinner, A-703 Wells Hall, Department of Linguistics and Languages, Michigan State University, East Lansing, MI 48824; e-mail: spinnerp@msu.edu.

© Cambridge University Press 2011

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Douglas, 1998; Read, 2007; Shohamy, 1994, 1998). The holy grail at the intersection of these two fields is the documentation of a clear path of grammatical development that could be used as part of an assessment tool.

Assessment experts rightly argue that language assessment tasks should include real language use, such as spontaneous or free production in conversation, rather than multiple-choice tests. However, it is not clear how to assess grammatical proficiency in spontaneous production. This difficulty has led to the use of holistic rubrics that provide only subjective descriptions of grammar use, such as "learners can *generally* [italics added] be understood by sympathetic interlocutors" versus "learners can *readily* [italics added] be understood by sympathetic interlocutors" (ACTFL, 1999, pp. 3–4). If concrete grammatical indicators of development could be included in rubrics like this one, the consistency and reliability of their use would be improved and they would have a stronger theoretical grounding. An assessment tool with concrete grammatical milestones would have myriad uses, both for research and in the classroom. For instance, such a tool could be used as a placement measure to ensure that the learners in each class are homogenous in terms of their developmental level or as a principled means of dividing learners into experimental groups in research studies.

There have been many obstacles to achieving this goal. First, determining a path of morphosyntactic development for learners of different language backgrounds has been difficult, and some might argue that it is an impossibility. Even if such a path could be determined, it is unclear whether a learner's stage on the path can be assessed in a straightforward and practical manner. An additional problem is that collaboration between the fields of language assessment and SLA has been limited because they have remained largely separate in their research agendas since the 1980s. Recent trends in the assessment of grammar in language assessment focus on the communicative effectiveness of learners—that is, what they can do with grammar—whereas SLA researchers are often more concerned with learners' competence—that is, what they know about grammar.

Despite these challenges, several proposals have been made in an effort to bridge the gap between language assessment and SLA with the creation of a morphosyntactic development measure for instructional or research purposes. The most well known is Pienemann's rapid profile (RP) measure, based on the stages of morphosyntactic development outlined in processability theory (PT; Pienemann, 1998, 2005). A RP is conducted by using targeted language tasks during an interview to elicit particular grammatical structures in the learner's speech; the presence or absence of these structures determines the learner's developmental stage. Rapid profile is intended to be fast and practical, taking less than 15 min. This is an important point because most efforts to place learners

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

along the PT scale have involved the collection of large amounts of data from learners. For instance, Dyson's (2009), Jansen's (2008), and Sakai's (2008) studies on PT all involved 45- to 60-min interviews with learners.

However, even though the RP is fast, there is one major practical limitation with using it for many pedagogical and experimental purposes: It requires a one-on-one interview with both an interlocutor and a trained observer, so that using it with large groups of learners—such as applicants to an English language program—would be unfeasible. Ellis (2008) has recently proposed an alternate measure—an oral imitation task—to place learners on the PT scale, but this too involves a one-on-one interview that would be impractical with large groups.

A second proposed developmental measure attempts to overcome this limitation: Young-Scholten, Ijuin, and Vainikka's (2005) organic grammar placement scale (OGPS) is based on Vainikka and Young-Scholten's (1994, 1996a, 1996b, 1998a, 1998b, 2006) minimal trees hypothesis, which proposes that learners acquire phrasal structures gradually. Like PT, organic grammar (OG) predicts stages of linguistic development that are characterized by the emergence of particular morphosyntactic elements in learners' grammar, and these stages are thought to hold for learners of all first-language (L1) backgrounds.

There are several practical advantages of the OGPS over the RP or Ellis's (2008) oral imitation task. First, either written or oral data can be used, so that the test could be easily administered with large groups (Young-Scholten & Ijuin, 2006a). Second, the OGPS includes a wider variety of grammatical forms, whereas the RP relies heavily on various types of questions, which often need to be elicited through targeted tasks. This difference is important because with the OGPS, spontaneous production data can be elicited through the use of prompts rather than in individual interviews with trained interlocutors, so that data can be collected from many learners at once by people without special training. However, the significant downside to the OGPS is that it remains largely untested as a measure of development for English, whereas PT, in contrast, has received support from numerous studies with a variety of L1s and L2s, including English (e.g., Di Biase & Kawaguchi, 2002; Glahn et al., 2001; Håkansson, Pienemann, & Sayehli, 2002; Kawaguchi, 2000; Sakai, 2008; but see Alhawary, 2009, for counterevidence).

In the current study, the feasibility of using the OGPS or RP scale (RPS) with large groups will be examined. This work is important because developmental scales are unlikely to be put to widespread use unless they are as convenient as current measures, such as writing samples and multiple-choice tests. The primary question is whether the two measures can accurately describe the order of acquisition of morphosyntax when used with easily obtained production data. If so, then the measures can contribute to a theoretically and empirically grounded assessment tool that is practical to use.

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

To answer this question, the current study examines small samples of oral production data from 48 adult learners of English-as-a-second-language (ESL) entering an intensive English program in the United States. As in previous research (e.g., Ellis, 2008; Meisel, Clahsen, & Pienemann, 1981), implicational scaling is used to determine whether the measures predict a developmental order in these data. The results of this analysis are also compared with the results from a traditional placement test used in the English program, the Michigan Test. The results indicate that the RPS successfully predicts the order of acquisition in the spontaneous production data. The OGPS may need to be further refined but includes important indicators of grammatical development. Elements of both scales could feasibly be included in the development of a new assessment tool.

MORPHOSYNTACTIC MEASURES

Rapid Profile

Pienemann's RP is a developmental measure based on profile analysis (Crystal, Fletcher, & Garman, 1976) and PT (Pienemann, 1998, 2005), which grew out of earlier work on the naturalistic acquisition of German as a L2 by child and adult immigrants (e.g., Meisel et al., 1981). PT posits a universal hierarchical order of language development based on processing procedures.

The stages of SLA in PT, which are predicted to hold for learners of all L1s and L2s, are (a) lemma access, (b) the category procedure, (c) the phrasal procedure, (d) the S-procedure, and (e) the subordinate clause procedure. The key factor in these stages is the structural distance over which feature matching—or unification in lexical functional grammar (Bresnan, 2001; Kaplan & Bresnan, 1982)—must take place.

In the first stage, lemma access, a particular word in the lexicon is activated. No feature unification can occur because no procedures are available yet. Only single words or short strings (memorized chunks) will be produced by learners at this stage, and the learner must rely largely on nonlinguistic strategies, such as gestures and facial expressions, for communication.

In the second stage, each lexical item can be associated with categorical information (noun, verb, etc.). Feature matching can only operate on elements within a single phrasal category because nothing else is available. For instance, learners will be able to produce plurals or possessives because the operations take place directly on the lexical item.

In the third stage, the category information stored with each word (e.g., noun) can serve as the head to a phrasal category (e.g., noun phrase [NP]). Now, feature matching can take place within a phrasal

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

category, such as plural agreement within a NP. For instance, consider the example in (1).

(1) [*these hats*]

In this phrase, the plural feature on *these* must match the plural feature on *hats*. This is feature matching within the NP.

Feature matching between phrases starts in the fourth stage. Most notably, this is the point at which subject-verb agreement becomes available. Consider the example in (2), in which feature matching takes place between the NP subject and the verb phrase (VP) predicate.

(2) [[$_{NP}$ *hats with wide brims*] [$_{VP}$ *protect you from the sun*]]

In (2), the plural feature on *hats* must match the plural feature on *protect*. The matching takes place between the two phrases.

In the final stage, a procedure is added that applies only to subordinate clauses. In English, it operates on *wh*-complement clauses, as shown in (3).

(3) a. *I wonder what they want.*
b. **I wonder what do they want.*

It should be noted that in English, these clauses use *wh*-words without the usual auxiliary inversion that takes place in questions. Because this procedure requires canceling a procedure acquired earlier, it is thought to be acquired particularly late. Being at the final stage implies that all the other stages have been passed through; that is, there is no way to miss a step or backtrack.

The specific English forms and structures that develop along with these processing procedures are spelled out in the RP table (Pienemann, 2003) shown in Table 1.

Note that the extended version of PT (Pienemann, 2005, 2007a) includes a number of other structures, including passives and topicalization. However, these structures have not been included in the RP developmental measure and for this reason they are excluded here.

The measure that Pienemann (e.g., 2003) advocated for assessing a learner's stage of development is the computer-based RP. During a RP assessment, which takes about 10–15 min, a learner and a native English-speaking interlocutor engage in spontaneous conversation that is intended to put the learner at ease. The observer then requests linguistic tasks, such as information-gap activities, that elicit specific forms (e.g., questions). A trained observer stationed at a computer listens for the relevant linguistic structures in the learner's speech and enters this information into the system.

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Table 1. Rapid profile stages

Stage	Phenomena	Examples
1	Words	*Hello, Five Dock, Central*
	Formulae	*How are you? What's your name? Where is X?*
2	S neg V(O)	*Me no live here/I don't live here.*
	SVO	*Me live here.*
	SVO?	*You live here?*
	–ed	*John played.*
	–ing	*Jane going.*
	Plural *–s* (noun)	*I like cats.*
	Possessive *–s* (noun)	*Pat's cat is fat.*
3	Do SV(O)?	*Do he live here?*
	Aux SV(O)?	*Can I go home?*
	Wh SV(O)?	*Where she went? What you want?*
	Adverb first	*Today he stay here.*
	Possessive (pronoun)	*I show you my garden. This is your pencil.*
	Object (pronoun)	*Mary called him.*
4	Copula S (x)	*Is she at home?*
	Wh-copula (x)	*Where is she?*
	V particle	*Turn it off!*
5	Neg /aux-2nd-?	*Why didn't you tell me? Why can't she come?*
	Aux-2nd-?	*Why did she eat that? What will you do?*
	3sg *–s*	*Peter likes bananas.*
6	Cancel aux-2nd	*I wonder what he wants.*

Note. The abbreviation SVO corresponds to subject-verb-object, neg to negative, aux to auxiliary, 3sg to third person singular.
Source. Adapted from "Rapid Profile," by M. Pienemann, 2003, http://groups.uni-paderborn.de/rapidprofile/docs/Stages.pdf

To determine that a learner has reached a certain stage, only the emergence of a form or structure is necessary because it indicates that a learner has the processing capacity required to produce it. According to Pienemann (1998), emergence is "(minimally) one occurrence in a sample" (p. 133). Note that using emergence as a measure means that learners' accuracy is not factored into the assessment. However, it is crucial that the emerged elements are productive rather than formulaic because memorized language can overrepresent learners' language ability (e.g., Myles, 2004). To determine productivity, Pienemann, Johnston, and Brindley (1988) advised eliminating from analysis any elements that demonstrate either of two characteristics. The first characteristic is the use of a particular form or structure with only a single lexical item—for instance, when a learner uses plural forms only with the word *years*, leaving all other nouns in the bare singular form. The second characteristic is the use of a structure in inappropriate contexts. For instance, a learner might use *don't know* to mean *can't* or *don't understand*.

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Pienemann (1998) suggested an additional measure of emergence for morphological markers such as plural *–s*: a distributional analysis to determine in which contexts the markers appear. However, for smaller samples such as the ones used in this study, there are often insufficient data for a distributional analysis. Therefore, for the current study, it will be necessary to adapt Pienemann's emergence criterion to small data samples.

Although elicited production data, such as the data elicited by the RP procedure, is generally used in PT research, Pienemann (1998) argued that any type of data should be sufficient as long as the same basic skill type is measured (the steadiness hypothesis). Indeed, research has shown that the developmental order of PT may be even more robust than this. For instance, Ellis (2008) demonstrated that data from a L2 oral imitation task reflects the PT order, and Pienemann (2007b) suggested that some kinds of written data may also prove to be adequate. However, not all types of data appear to produce the developmental order. For example, Ellis showed that a grammaticality judgment task failed to place learners along the PT developmental continuum, presumably because it reflects explicit knowledge of grammar.

One of the goals of the current study is to determine whether even short samples of spontaneous production data reflect the PT developmental continuum. Most studies of PT use interviews of 45–60 min; even RP takes 10–15 min one-on-one with an interlocutor and an observer. However, short samples of spontaneous production data that can be collected with prompts are easily elicited and, therefore, practical to obtain from large groups, as Young-Scholten et al. (2005) indicated. It is therefore highly desirable to find a morphosyntactic measure that could feasibly be used with these data. However, the significant disadvantage of using short samples is that a large number of the elements listed in the RP table will not be available—particularly questions. One of the aims of this study is to determine the usefulness of applying RP in a limited way to these kinds of data.

Organic Grammar

Vainikka and Young-Scholten's (2006) OG is based in generative syntactic theory and assumes the operation of Universal Grammar in SLA. Specifically, OG proposes a L2 version of the weak continuity hypothesis (e.g., Clahsen, 1990/1991, Clahsen & Penke, 1992; Vainikka, 1993/1994), which claims that L1 learners acquire grammatical phrases gradually, building a syntactic tree from the ground up. According to OG, L2 learners begin with a VP, adjust the headedness of the phrase (if necessary), and then move on to a functional phrase (FP), an inflectional phrase (IP), and,

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

finally, a complementizer phrase (CP). Thus, there are five main stages. It is crucial to note that each of these stages is claimed to be characterized by the presence (and absence) of morphosyntactic elements that correspond to the available phrases. Therefore, learners' stages of development can be determined by analyzing the morphosyntactic elements that they produce.[1]

As in PT, the progression in OG is based largely on the emergence of elements, not on accuracy. Young-Scholten et al.'s (2005) criteria for determining the productivity of emerging elements are similar to those of Pienemann (1998); that is, the authors recommend ascertaining that a structure is used with several different lexical items. Also similar to Pienemann's recommendations, the procedure for dealing with morphological markers is somewhat more involved. For morphology, Young-Scholten et al. recommend determining productivity based on suppliance in obligatory contexts; that is, a form must be used in 60% of required contexts to be considered emerged. The purpose of this rule is to determine that a form is being used appropriately and not in a random fashion.

Table 2 lists the various elements that are associated with each phrase (and, therefore, stage). Between the VP and the IP stages, learners develop a functional projection (the FP) above the VP. This projection allows for some morphosyntactic elements associated with the IP stage to appear, even though the IP stage has not yet emerged. For instance, learners at the FP stage may produce modals, auxiliaries, and some sentences with an initial adverb, but they will not yet have inflection on verbs.

As with PT, the influence of the L1 on the development of these elements is predicted to be minimal. In fact, according to OG, the L1 will only affect the morphosyntax of the very first stage of learning—the initial VP stage. At the initial VP stage, learners will produce verbs and complements that reflect the order in their L1 (either object-verb or verb-object). After this stage, all learners will correct the word order (if necessary) and move on through the other stages in roughly the same way, regardless of the L1.[2] Vainikka and Young-Scholten have provided

Table 2. Morphosyntactic elements associated with phrases

Phrase	Morphosyntactic elements
VP	Verbs and complements, word order
IP	Subjects, pronouns, *do*-support (as in negation), copula, modals, sentences with expletive subjects (*there* and *it*), tense, aspect, and agreement marking, sentences with a noninitial subject
CP	Complement/noun clauses with *that*, adverb (adjunct) clauses, relative clauses, *wh*-questions, passive

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

evidence for these claims and articulated these ideas more fully in a series of articles (Vainikka & Young-Scholten, 1994, 1996a, 1996b, 1998a, 1998b, 2006; Young-Scholten & Ijuin, 2006a, 2006b). However, for reasons of space, the discussion now turns to the assessment measure, the OGPS.

To evaluate L2 learners based on the OGPS, it is necessary to elicit the relevant morphosyntactic structures. As with PT, any type of spontaneous production data should be sufficient as long as the data are not overly influenced by metalinguistic knowledge that may not reflect learners' true linguistic level. Generally speaking, oral data are considered to be less influenced by metalinguistic knowledge because they allow less time for planning and correction. However, Young-Scholten et al. (2005) wanted to have the option of using written data, which are easier to collect from large groups. To avoid the influence of metalinguistic knowledge and thus to mimic oral production as much as possible, Young-Scholten et al. suggested introducing a time constraint and concealing the true nature of the written OG task by describing it as a memory test. In their study, they asked 44 ESL students of various L1 backgrounds to write brief descriptions of slides of various objects, events, and people. The learners produced fairly short writing samples, from 48–195 words.

These data were then examined to determine the presence or absence of the forms in Table 3. Generally speaking, these stages correspond to the stages in minimal trees theory. However, note that Stage 4 involves simple subordination and Stage 5 involves complex subordination. Young-Scholten et al. (2005) explained that the difference involves displacement; for instance, subject relative clauses are simpler than other relative clauses because there is less displacement. However, the difference between simple and complex subordination, or simple and complex aspect, is not part of the minimal trees hypothesis nor is it fully articulated in Young-Scholten et al. Therefore, the current study uses only those elements that are detailed in Table 3 and makes no attempt to apply other indicators of simple versus complex language use. In the future, providing complete theoretical description for the distinction between simple and complex language may prove to be very useful for the OGPS.

In their 2005 study, Young-Scholten et al. argued that OG placed students in a manner similar to the more traditional method that had been used—a reading task and writing sample. However, where the two measures differed, OG placed students more appropriately (based on the teachers' classroom assessments). The current study also compares the RPS and OGPS with traditional placement measures.

It is worth noting that there are significant differences between the RPS and OGPS. For instance, a number of features are included in the RPS but not OGPS, such as object pronouns, possessive pronouns and

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Table 3. Organic grammar placement scale

Stage	Word order in declaratives	Types of verbs	Verbal agreement and tense marking	Pronouns	Complex syntax
1	Initially resembles the L1	Thematic (main) verbs only	None	Pronouns absent	None
2	Resembles the L1	Thematic verbs, copula *is* appears	None	Pronoun forms begin to emerge	Formulaic or intonation-based questions
3	Resembles the L2	Thematic verbs, modals, copula forms beyond *is*	No agreement, some tense, some aspect, but not productive	More pronouns, but they can still be missing	Questions formulaic or without inversion, conjoined clauses
4	Resembles the L2	Thematic verbs, modals, copula forms beyond *is*, range of auxiliaries emerges	Productive tense (60% SOC), aspect, agreement with *be* forms	Pronouns obligatory, *there* and existential *it* emerge	Productive questions (but may still lack inversion), simple subordination
5	Resembles the L2	Complex tense, aspect forms; passives; range of thematic verbs, modals, auxiliaries	Forms usually correct, apart from those newly attempted	Use of *there* and *it* beyond stock phrases	All questions with inversion, complex subordination

Note. SOC = suppliance in obligatory context.
Source. Adapted from "Organic Grammar as a Measurement of Development," by M. Young-Scholten, C. Ijuin, & A. Vainikka, 2005, unpublished manuscript, p. 9.

nouns, and *wh*-complement clause structures. Similarly, the OGPS includes a number of forms and structures not included in the RPS: modals, relative clauses, adverb clauses, and expletive subjects, to name a few.

THE PRESENT STUDY

In this study, the features of both the RPS and OGPS that appear in short samples of spontaneous production data are examined. The overall purpose is to determine whether these elements emerge predictably and can, therefore, contribute to a morphosyntactic assessment measure that requires only a small data sample elicited with prompts rather than the interviews that have previously been used for morphosyntactic assessments for PT. The specific research questions are as follows:

1. Does the path of development laid out in the RPS predict the order of emergence of morphosyntactic features for students entering an intensive English program when a short sample of spontaneous production data is used?
2. Does the path of development laid out in the OGPS predict the order of emergence of morphosyntactic features for students entering an intensive English program when a short sample of spontaneous production data is used?

If the elements in the RPS and OGPS emerge in a predictable way in the short samples of production data, then those features might be used to create a measure of morphosyntactic development that could be used with large groups.

To explore how such a measure might differ from current placement measures, a third research question was also addressed:

3. How do the assessment outcomes based on the OGPS and RPS with short samples of spontaneous production data compare to the assessment outcomes of a standard placement or proficiency test (i.e., the Michigan Test)?

Data

The current study took place at the English Language Institute, an intensive English program that offers noncredit courses at four levels: preintermediate (Level 2), low intermediate (Level 3), high intermediate (Level 4), and advanced (Level 5). There are no classes for beginners. Students are placed into class levels based on their performance on three measures: the Michigan Test of English Language Proficiency

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

(MTELP), the University of Michigan Listening Comprehension Test (LCT), and a writing sample. The MTELP and LCT are traditional multiple-choice tests that measure grammar, vocabulary, and reading and listening comprehension. The writing sample is graded by English Language Institute instructors using a rubric that measures content, organization, and grammatical accuracy. Table 4 shows the scores required to be placed into each of the four proficiency levels.

As a measure of spontaneous oral production, this study uses recorded speaking activities (RSAs), an oral assessment tool developed by the English Language Institute that is used in speaking classes and that could be adapted for placement purposes. Currently, RSAs are administered four times per semester in every proficiency level. During a RSA, each student sits at a computer. The instructor announces a question; the students then have a full 2 min to discuss their answer without notes or references. The students record their speech on a Sony HS 90-B microphone headset, using an interface built into the software Runtime Revolution. When the 2 min have passed, the computer automatically stops recording.

Topics for each semester are created by supervisors at the institute. The questions are designed to be appropriate for students of any age, class level, and cultural background; they are also broad enough that students will not find it difficult to speak for the full 2 min. At least one RSA per semester is intended to elicit past tense; otherwise, the questions are not explicitly designed to elicit particular structures. A sample question is given in (4):

(4) *Describe something that you liked to do when you were in your country but that you can't do here. Where did you do this? Why did you like it? How did it make you feel?*

A full list of the RSA questions used in this study is given in the appendix. All class levels answered the same questions each semester, and no student in the study was given a particular question more than once.[3] There are three or four samples of data per semester. For each

Table 4. Guidelines for new student placement at the English Language Institute

Level	MTELP	LCT	Writing sample
Not accepted	0–32	0–32	0
2	33–44	33–44	1
3	45–59	45–59	2
4	60–79	60–79	3
5	80–100	80–100	4–5

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

student, the samples from each semester have been collapsed into a single large file.[4]

Participants

Data from 48 of the participants entering the English Language Institute were analyzed. Participants were selected from four L1 groups: Arabic, Korean, Chinese, and Spanish. These languages were selected for two reasons: First, these were the largest groups in the institute and, second, the languages have diverse morphosyntactic systems (recall that OG and PT predict little L1 transfer). It was not possible to have a balanced number of students from each language group due to the distribution of students in the institute; however, as many as possible were included. Additionally, it was not possible to include a balanced number of students who tested into each of the four levels (using the results of the Michigan Test) because only one learner from the selected L1 backgrounds (a Spanish speaker) tested into Level 5. Therefore, so that there would be at least a few samples of advanced speech, a Thai speaker and a Japanese speaker who tested into Level 5 were also included. Table 5 shows the distribution of participants by L1 and by the level that they tested into by means of the traditional placement measures.

Participants' ages ranged from 18 to 53 (M = 25.7, SD = 6.1). Thirteen had studied languages other than English, and nearly all reported studying English for at least 5 years prior to placement. The majority had spent less than 2 weeks in the United States (i.e., the time between arriving in the country and beginning study in the English Language Institute).

Each student was assigned a participant number, which will be used in the reporting of the results. The letter that precedes each student's number indicates the student's L1 (A = Arabic, K = Korean, C = Chinese,

Table 5. Distribution of participants by L1 and level

L1	Level 2	Level 3	Level 4	Level 5	Total
Arabic	4	6	4	—	14
Korean	5	5	4	—	14
Chinese	2	5	4	—	11
Spanish	2	2	2	1	7
Japanese	—	—	—	1	1
Thai	—	—	—	1	1
Total	13	18	14	3	48

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

S = Spanish, J = Japanese, T = Thai). The final number indicates the level into which the student was placed by traditional means. For example, A45-2 indicates that student #45 is an Arabic speaker who tested into Level 2 upon arrival at the institute.

Measures

The primary measure in this study is a count of specific morphosyntactic elements in learners' spontaneous oral production. The morphosyntactic elements chosen for the study are those that are predicted by Young-Scholten et al. (2005) to emerge with the building of tree structure as well as the features predicted by PT to emerge with processing development in English, as listed in the RPS (Pienemann, 2003). Both OG and RP rely on the emergence of forms and structures rather than the accuracy of their use. A form or structure is generally considered to have emerged at its first productive use, but determining productivity is often difficult. In this study, several steps were taken to eliminate formulaic uses. First, any idiomatic expressions or utterances that repeated part or all of the topic prompt were eliminated. Second, any forms that appeared to be formulaic following Pienemann et al.'s (1988) criteria were excluded. These eliminations amounted to about 0.7% of the utterances collected.

Many researchers advocate finding multiple tokens of a particular element with different lexical items before determining that it is productive. In short data samples—such as the ones used here or in Young-Scholten et al.'s (2005) study—the emergence criteria has to be adapted somewhat, as it is often not possible to find multiple forms for certain elements, such as relative clauses. Therefore, any time a learner's result relies on a single token (e.g., a single modal verb or a single relative clause), that result should be considered tentative; in the current study, this is shown in the results as an asterisk between dashes (-*-). When there are at least two (different) tokens on which to base the analysis, a plain asterisk is shown, and when a large number of widely varying tokens are present, two asterisks are used. This system is used to provide some additional information beyond a simple yes/no measure.

Implicational Scaling

Implicational scaling indicates whether a true developmental pattern is present in data (Meisel et al., 1981). To create the scale, all counts of morphosyntactic elements were entered into two tables in Microsoft

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Excel—one for the OGPS and one for the RPS. The elements predicted to appear earlier are on the left, and the elements predicted to appear later are on the right. For instance, the use of correct subject-verb-object (SVO) word order is one of the first predicted milestones for a L2 learner in the OGPS, and so the column with those results is the leftmost one. Given that adverb clauses are predicted to appear later, that column is on the far right. Each asterisk, double asterisk, or asterisk in dashes is counted as 1 point, and the total points are tallied for each learner. The data are then sorted by number of points per learner, so that the learners with the lowest number of points appear at the top of the table and those with the highest number of points appear at the bottom.

Learners with the same point tally may be rearranged so as to reduce the number of errors in the table (Hatch & Farhady, 1982) because there is assumed to be development within each stage, and learners may thus be at different points in these within-stage continua. Similarly, the elements that are predicted to occur in the same stage of learning may be rearranged so as to have the best fit. For instance, in the RPS, *–ed, –ing*, and plural *–s* are all predicted to emerge in Stage 2. These three elements may be rearranged within that stage so as to achieve the best table.

A line is then drawn on the table to separate the emerged side from the not yet emerged side. This would ideally be a diagonal line from the upper left corner to the bottom right corner. An error occurs in the table when there is an N (which indicates a nonemerged element) on the emerged side or an asterisk on the not-emerged side of the line. For instance, in (idealized) Table 6, student #5 has an N out of place and an asterisk out of place; these represent errors in the table, where predictions were not met.

There are two simple calculations that are used to determine whether a set of data makes a true developmental pattern. The first is the coefficient of reproducibility (C of R), which is based on the number of errors in the table and which indicates how predictable the results are for all

Table 6. An implicational table with two errors

Student	A	B	C	D	E	F	G	Points
1	N	N	N	N	N	N	N	0
2	*	N	N	N	N	N	N	1
3	*	*	N	N	N	N	N	2
4	*	*	*	N	N	N	N	3
5	*	N	*	*	N	*	N	4
6	*	*	*	*	*	N	N	5
7	*	*	*	*	*	*	N	6
8	*	*	*	*	*	*	*	7
Points	7	5	5	4	3	3	1	

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

the individuals. A C of R over .9 is generally considered to indicate a predictable pattern (Hatch & Farhady, 1982). The C of R can also be calculated for individual items. This figure provides information about how predictable the emergence of a single element is. The second calculation is the coefficient of scalability (C of S). This figure is intended to indicate if a set of variables is truly scalable (i.e., a developmental pattern). The coefficient should be over .6 or .65 for the table to be considered scalable (Hatch & Farhady, 1982).

One difficulty with implicational scaling occurs with spontaneous production data (such as those used in this study). In spontaneous data, there may not be context for certain elements. For instance, a learner may not happen to use any third-person agreement during a given recording. If data are missing, the researcher must decide what to do with those squares in the table. In this study, they are treated as inert; that is, they are not included in any counts—either as errors or correct responses, or in the total count of responses. However, if necessary, they can be used to move a learner into a more appropriate developmental group. For example, a learner with a total point count of 4 but with one cell of missing data may be moved into the group of learners with a count of 5 (but not 3). This adjustment ensures the best possible fit for the table.

Hatch and Farhady (1982) pointed out another issue with spontaneous production data that is particularly important for the OGPS—that is, calculating suppliance in obligatory context when there are only a few tokens of data. For instance, if a learner only has two contexts for the plural, and one of them is marked incorrectly (i.e., the singular form is used), should the researcher eliminate the data completely or treat it as 50% suppliance? For this study, the minimum number of occurrences has arbitrarily been set at four tokens of a particular form when calculating suppliance in obligatory contexts. For instance, if a learner only uses three past tense verbs, those data are eliminated, as if the learner had not had any context for past tense at all. This practice is intended to ensure that a reasonable number of tokens are used to calculate percentage of suppliance. Note that four occurrences are only required when dealing with percent measures (tense and agreement in OGPS); all other measures are based on emergence—one productive form is enough to tentatively consider the element acquired.

RESULTS

Results for the RPS

The main difficulty with using the RPS with the short samples of production data was that only a small portion of the elements could be included. Particularly problematic was the exclusion of question data, which is

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

crucial for distinguishing Stage 4. Despite this difficulty, the RPS was successful in that it accurately described the order of acquisition for these learners. The implicational table with the elements that could be included is shown in Table 7.

Table 7. Implicational table: Predictions of RP

Student	WO	Plur	*–ing*	*–ed*	Poss pro	Adv 1st	Obj pro	3rd *–s*	Cncl aux
A148-2	*	N	-*-	/	*	N	*	N	N
A25-2	*	*	N	-*-	*	N	N	/	N
K123-2	*	*	N	*	*	N	N	/	N
K266-2	*	*	*	*	*	N	N	/	N
C126-2	*	*	N	/	N	*	N	/	N
A170-2	*	*	N	*	*	*	N	N	N
S366-2	*	*	N	-*-	*	*	N	/	N
S362-2	*	*	N	*	*	*	N	/	N
A45-2	*	*	-*-	-*-	N	*	N	/	N
A159-3	*	*	*	*	*	*	N	N	N
A338-4	*	*	*	/	*	*	N	*	N
A199-4	*	*	*	*	*	*	N	/	N
K111-3	*	*	-*-	*	*	*	N	/	N
T397-5	*	*	*	*	*	*	N	/	N
K101-3	*	*	*	*	*	*	N	-*-	N
J84-4	*	*	*	*	*	*	N	-*-	N
K167-3	*	N	N	*	*	*	-*-	/	N
K156-2	*	*	N	-*-	*	*	*	/	N
A12-2	*	*	*	-*-	N	*	-*-	/	N
K267-2	*	*	*	N	*	*	-*-	/	N
K300-3	*	*	N	*	*	*	*	/	N
C282-3	*	*	N	*	*	*	*	N	N
C298-3	*	*	-*-	N	*	*	*	N	N
S100-5	*	*	N	*	*	*	*	/	N
C271-2	*	*	N	N	*	*	*	-*-	N
A163-3	*	*	*	*	*	*	*	N	N
A279-4	*	*	*	*	*	*	*	/	N
K269-2	*	*	*	*	*	*	*	/	N
K46-3	*	*	-*-	*	*	*	-*-	/	N
K217-4	*	*	N	*	*	*	*	/	*
K276-4	*	*	*	*	*	*	*	N	N
C127-3	*	*	-*-	*	*	*	-*-	N	N
C201-4	*	*	*	*	*	*	*	/	N
S391-4	*	*	*	/	*	*	-*-	N	*
S130-3	*	*	*	/	*	*	*	*	N
A160-3	*	*	*	-*-	*	*	*	*	N
A161-3	*	*	*	*	*	*	*	*	N

continued

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Table 7. Continued

Student	WO	Plur	*–ing*	*–ed*	Poss pro	Adv 1st	Obj pro	3rd *–s*	Cncl aux
A182-4	*	*	*	*	*	*	*	-*-	N
K320-4	*	*	-*-	*	*	*	*	-*-	N
C177-3	*	*	*	*	*	*	*	*	N
C301-3	*	*	-*-	*	*	*	*	-*-	N
C118-4	*	*	*	*	*	*	-*-	*	N
S173-3	*	*	*	-*-	*	*	*	*	N
C278-4	*	*	*	N	*	*	*	*	*
A157-3	*	*	*	*	*	*	N	*	*
S383-4	*	*	*	/	*	*	-*-	*	*
J274-5	*	*	*	*	*	*	*	/	*
K275-4	*	*	*	*	*	*	*	-*-	*

Note. WO = word order; Plur = regular plural (*–s*); *–ing* = verbs with *–ing*; *–ed* = regular past-tense-marked verbs; Poss pro = possessive pronouns; Adv 1st = adverb (or prepositional phrase, etc.) as first element in a sentence; Obj pro = object pronouns; 3rd *–s* = third-person singular inflection(s); Cncl aux = *wh*-complement clauses; *, **, or -*- = emerged feature; N = not emerged; / = no context for this feature.

There are 30 errors (Ns on the left and *s on the right) in the table out of 404 total cells (excluding 28 nonapplicable responses; i.e., where there was no context for inflection). Therefore, the C of R is .93 [1 – (30/404)]. This indicates that the results for individuals can be predicted based on the table. Additionally, the C of S is .7 [.16/(1 – .77)]. This number is greater than .6–.65 and, therefore, indicates a scalable table (i.e., a real developmental pattern). If the table were to be reordered, the table would not produce a higher C of R or C of S; in other words, this order of acquisition best describes these data.

The individual Cs of R have been calculated for the elements in the table; the results are given in Table 8.

Nearly all of the elements reach .9 or come very close when the RPS is used with these data. The only element that fails to meet the criterion is *–ing*, which does not appear in a predictable way in these data.

Note that when using the small number of elements included in these short samples, the line separating the emerged from the not emerged side is on the far right; that is, the majority of the elements are emerged for most learners. In fact, one of the structures—correct SVO word order—emerged for all learners. The result is that the elements appear to distinguish the beginning levels best. By the time students have reached intermediate to advanced levels, nearly all of the elements on the table have emerged in their production data.

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Table 8. Coefficients of reproducibility for elements in the RP table

Element	C of R
Plural	.96[a]
–ing	.73
Past	.90[a]
Possessive pronoun	.93[a]
Adverb 1st	1.00[a]
Object pronoun	.95[a]
3rd singular *–s*	.89[b]
Cancel aux 2nd	.95[a]

[a] A predictable pattern (Hatch & Farhady, 1982).
[b] Approaches the .9 standard for predictability.

OG with Short Samples of Spontaneous Production Data

An implicational table (Table 9) was created for the OGPS measures, which include certain morphosyntactic elements that are somewhat different than those in the RPS.

Notice that the line separating the emerged from the not-emerged side is drawn high and far to the right in the table, so that the majority of the table is emerged (with the exception of passive), similar to Table 7, based on RPS. This line was drawn in this way because it created the minimum amount of error.

There are 94 errors in Table 9 (Ns on the left and *s on the right of the line) out of 682 total cells (excluding 38 nonapplicable responses). Therefore, the C of R is .86 [1 – (94/682)]. This result is lower than is required to indicate a predictable table (although not by a great deal). In other words, the results of individual students cannot be reliably predicted based on this order of acquisition. The C of S is .42 [.1/(1 – .76)]. This number is lower than the .6–.65 that indicates a scalable table (i.e., a real developmental pattern).

It is also possible to test the reproducibility of individual test items. This information could be particularly useful here to discover which elements are causing the table to fail. The C of R was calculated for each of the items in Table 9; these data are presented in Table 10. Results for word order, subjects, and pronouns are not included because those elements had emerged for every student. Similarly, the use of the copula and coordination score highly by this measure, but this fact is only trivially true, given that only one learner does not yet have coordination and only one lacks copula forms beyond *is*.

Given that a number of learners fail to produce modals and passive, it is more interesting to see that these two elements also have high Cs of R

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Table 9. Implicational table: Predictions of OGPS

Student	WO	Pro	Sub	Cop	Co	Mod	Pst	Agr	Asp	B/c	RC	Ex.	CC	AC	Pv
A25-2	-*-	*	*	*	-*-	N	/	/	N	*	N	N	-*-	N	N
A148-2	*	*	*	-*-	*	N	*	N	N	*	N	N	N	N	N
K123-2	*	*	*	*	*	N	*	/	N	N	N	N	*	N	N
C126-2	*	*	*	-*-	-*-	N	/	/	N	-*-	N	-*-	-*-	N	N
A12-2	*	*	*	N	*	N	*	/	N	N	N	N	-*-	-*-	N
K267-2	*	*	*	-*-	*	N	*	/	N	N	N	N	-*-	N	N
K269-2	*	*	*	*	*	N	*	/	-*-	*	N	N	N	-*-	N
K266-2	*	*	*	**	-*-	-*-	N	/	N	-*-	N	N	N	-*-	N
K167-3	*	*	*	**	*	-*-	N	/	N	*	N	N	*	-*-	N
K156-2	*	-*-	-*-	**	-*-	-*-	*	/	N	N	N	N	-*-	-*-	N
C298-3	*	*	*	*	*	**	/	N	N	N	N	-*-	*	*	N
C271-2	*	*	*	**	*	-*-	*	/	N	-*-	N	N	*	*	N
S366-2	*	*	*	*	-*-	*	/	/	N	-*-	*	**	**	N	N
A45-2	*	*	*	*	*	-*-	*	/	N	*	-*-	*	-*-	N	N
C127-3	*	*	*	**	*	*	*	-*-	N	*	N	-*-	**	N	N
S173-3	*	-*-	-*-	**	*	*	N	N	-*-	*	N	N	*	**	N
A338-4	*	*	*	**	*	**	/	N	N	*	**	N	**	*	N
K300-3	*	*	*	*	*	-*-	*	/	N	*	N	N	-*-	*	N
A182-4	*	*	*	*	*	*	N	-*-	N	N	*	N	**	**	-*-
K217-4	*	*	*	**	*	**	*	/	N	N	N	*	**	**	N
A170-3	*	*	*	**	*	**	*	N	N	-*-	-*-	-*-	N	*	N
K111-3	*	*	*	**	*	*	*	/	N	*	N	*	-*-	*	N
A159-3	*	*	*	*	*	*	*	N	-*-	-*-	N	N	-*-	-*-	N
K46-3	*	*	*	**	*	N	*	/	-*-	-*-	N	-*-	*	*	N
C282-3	*	*	*	**	*	-*-	*	/	N	*	*	N	**	-*-	-*-
S391-4	*	-*-	-*-	**	*	**	/	/	*	*	**	N	**	*	N
K101-3	*	*	*	**	*	**	*	/	N	*	N	*	*	*	N
A279-4	*	*	*	**	*	**	*	/	N	*	*	N	**	*	N
A157-3	*	*	*	**	*	*	*	/	N	*	**	*	*	-*-	N
S130-3	*	-*-	-*-	**	*	**	/	-*-	*	*	-*-	N	**	-*-	N
S362-2	*	*	*	*	*	**	*	/	N	*	-*-	*	**	-*-	N
T397-5	*	*	*	**	*	*	*	/	-*-	*	*	*	**	N	-*-
C177-3	*	*	*	**	*	*	*	*	N	*	N	-*-	**	*	N
C84-4	*	*	*	*	*	**	*	/	N	*	**	-*-	**	**	N
C278-4	*	*	*	*	*	**	N	N	-*-	*	*	-*-	**	**	-*-
C201-4	*	*	*	**	*	*	*	/	*	*	-*-	N	**	*	N
C301-3	*	*	*	**	*	**	N	-*-	N	*	*	*	**	**	N
K320-4	*	*	*	**	*	**	N	/	*	*	*	-*-	**	**	N
C118-4	*	*	*	**	N	**	*	*	N	*	N	**	**	*	-*-
S383-4	*	-*-	-*-	**	*	**	/	*	**	*	*	*	**	-*-	N
A160-3	*	*	*	*	*	*	*	*	-*-	*	-*-	N	-*-	*	N
A163-3	*	*	*	**	*	-*-	*	N	**	*	-*-	-*-	**	*	N
K275-4	*	*	*	*	*	*	*	/	-*-	-*-	-*-	**	**	**	N

continued

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Table 9. Continued

Student	WO	Pro	Sub	Cop	Co	Mod	Pst	Agr	Asp	B/c	RC	Ex.	CC	AC	Pv
A199-4	*	*	*	**	*	-*-	*	/	*	*	-*-	*	**	**	N
A161-3	*	*	*	**	*	*	*	-*-	*	*	-*-	-*-	**	**	N
K276-4	*	*	*	**	*	**	*	-*-	-*-	*	N	N	**	**	-*-
S100-5	*	*	*	**	*	**	*	/	-*-	*	**	*	**	**	**
J274-5	*	*	*	**	*	**	*	/	N	*	**	-*-	**	**	-*-

Note. WO = word order (SVO); Pro = no missing pronouns; Sub = no missing subjects; Cop = copula forms other than *is*; Co = coordination; Mod = modals; Pst = 60% past tense marking; Agr = 60% agreement on third-person singular; Asp = progressive or perfect aspect; B/c = clauses with *because*; RC = relative clauses; AC = adverb clauses; CC = complement clauses; Ex = expletive subjects; Pv = passive.

(.98 modals, .92 passive). That means that learners' use of modals and the passive is highly predicted by Table 10. However, none of the other elements emerge in a predictable fashion based on this table. Note that the low coefficients do not indicate that the elements never emerge in a predictable way; they only indicate that they do not emerge predictably based on this order of acquisition and with these data and methodology.

Notice that, as with the RPS table (Table 7), the students who tested into Level 2 by the traditional measure cluster at the top of Table 10; that is, they predictably fail to produce many of the morphosyntactic forms analyzed here. However, the students who tested into Levels 3, 4, and 5 are scattered, and their ordering is quite different from the or-

Table 10. Coefficients of reproducibility for elements in OGPS

Element	C of R
Copula	.98[a]
Coordination	.98[a]
Complement clause with *that*	.79
Past tense	.75
Because clause	.81
Agreement	.67
Modals	.98[a]
Adverb clause	.83
Expletives	.73
Relative clause	.77
Aspect	.58
Passive	.92[a]

[a] A predictable pattern (Hatch & Farhady, 1982).

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

dering in the RPS table; that is, the students that score the highest on the OGPS are not the same students who score the highest on RPS. These observations merit further exploration.

DISCUSSION

Research Questions

Returning now to the research questions posed earlier, the first question asked whether the path of development laid out in the RPS could predict the order of emergence of morphosyntactic features for students entering an intensive English program when a short sample of spontaneous production data is used. An implicational table based on the RPS with these data was both reproducible and scalable; that is, word order, plural, regular past tense marking, and possessive pronouns predictably emerge before sentences beginning with a nonsubject, object pronouns, third-person singular marking, and *wh*-complement phrases. These elements emerge predictably despite the limited length of each learner's data, which provided little context for many elements. This study therefore provides evidence that PT can apply to a variety of data types. This finding is encouraging because it also raises the possibility of including these morphosyntactic elements in a developmental measure for short samples of spontaneous production data. The use of *–ing* forms is the only element that is not reproducible in this table.

The second research question asked whether the path of development laid out in the OGPS could predict the order of emergence of morphosyntactic features for students entering an intensive English program when a short sample of spontaneous production data is used. Some predictions of the OGPS appear to be borne out. For instance, it is true that for nearly all learners, word order, subjects, coordination, and pronouns emerge early, whereas the passive emerges late. Clauses with *because* generally appear before other types of adverb clauses. Additionally, modals appear to be a particularly useful predictor of development in that they emerge in a highly predictable fashion, dividing low-level learners from intermediate learners. However, for the most part, the morphosyntactic development of the learners in this study could not be adequately described by the OGPS—at least not with this methodology. Implicational scaling demonstrated that the morphosyntactic elements were not acquired in the predicted developmental order, which assumes a structure building model in which IP features are acquired before CP features. For example, there are a number of learners who fail to produce modals (a feature of Stage 1 and Stage 2) but who produce elements that are predicted to appear in Stage 4 or Stage 5: complement

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

clauses with *that*, adverb clauses, and productive past tense marking. Note, however, that because there are far more elements of the OGPS to be found in this dataset, it is not surprising that there are a higher number of errors than for the RPS.

A few specific elements are consistently acquired earlier or later than predicted. Let us consider several aspects of the OGPS that seem particularly problematic. First, adverb clauses appear to be produced much earlier than predicted. For example, learner C271-2 produces only one modal verb and no relative clauses existential, or past tense marking on main verbs. By the predictions of the OGPS, this participant should not be able to produce Stage 5 structures such as adverb clauses, and yet several appear in her data, shown in (5) and (6).[5]

(5) ***When I was high school student,*** *I was 14 years old.*
(6) ***If I have time,*** *I want to back to Taiwan to see him and tell him my new life.*

Another element that appears out of place in the developmental order is aspect marking, which is particularly late in emerging. The OGPS predicts that it emerges simultaneously with tense, but these data show a different pattern. A large number of learners who show productive tense (measured at 60% suppliance) do not produce progressive or perfect aspectual marking at all. For instance, consider learner K46-3, who produces appropriate past tense marking (both regular and irregular), as shown in (7).

(7) *The first time I* ***arrived*** *there, I* ***boarded*** *a kind of ship…I remember that when I* ***boarded*** *the ship, I* ***saw*** *blue sky and clean lake and also I* ***could*** *meet another country friends, so I* ***thought****, um. When I* ***traveled*** *to Switzerland I* ***couldn't*** *speak English very well…*

However, this learner fails to use any aspectual marking, even when it is required by context. An example is given in (8); note that either progressive or perfect aspect is required on the verb *change*, but the learner uses simple (presumably default) aspect.

(8) *When I was a freshman, there weren't many kinds of scholarship in my university. But now, many things* ***change*** *in my university.*

Another type of aspectual error is also possible: that of producing perfect or progressive aspect in inappropriate contexts. For instance, C84-4, who has 82% suppliance of past tense (and is therefore considered to have emerged past tense), overuses pluperfect in the example in (9).

(9) *I* ***had gone*** *to Hawaii for ten years ago.*
"I went to Hawai'i ten years ago."

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Many other learners overuse progressive, as shown by the example in (10).

(10) *My grandfather of my father is an important person for me because he* ***was taking*** *care of us, of all my family in his era and he had a job in the fire station, and he* ***was earning*** *a high salary which* ***is making*** *a lot of money to build our house*. (A45-4)

It is difficult to know what to make of these cases. On the one hand, it could be claimed that these learners have the appropriate morphosyntax in place to produce these features; that is, in order to produce the auxiliary verbs with participles, they must have IP and, therefore, they can be considered to have emerged aspect. On the other hand, this seems a problematic way to measure proficiency if both correct and incorrect uses are considered equivalent.

Young-Scholten et al. (2005) suggested some ways to further articulate these issues. They suggested that it may be possible to take some accuracy issues into account, or at least examine the variety of forms that a learner is able to produce in a given category (e.g., different aspectual forms). Exploring these issues may make the OGPS predictions more accurate.

The last research question asked how the assessment outcomes based on the OGPS and RPS with short samples of spontaneous production data compare to the assessment outcomes of a standard placement or proficiency test (i.e., the Michigan Test). The OGPS, RPS, and the standard placement test produce the same results for most of the Level 2 (preintermediate) students. However, for intermediate levels, the three tests produce different results. Note that if the developmental measures produced the same result as the traditional placement measures, in the implicational tables, the Level 2 students would be followed by the Level 3 students, who would be followed by the Level 4 students, and so on. There are several possible explanations for the fact that there is such a large discrepancy between the predictions of the standard placement test and the morphosyntactic measures. First, it could mean that morphosyntactic development does not progress in parallel with increasing proficiency but rather takes place quite independently (at least at advanced levels). This would not be an entirely unexpected finding, given that the Michigan Test takes many factors into account (such as vocabulary and listening skills) that are not included in the OGPS and RPS. Note, however, that Young-Scholten et al. (2005) found that their study using OGPS produced results very similar to a standard placement test.

There is a second issue worth mentioning here. Using the OGPS and RPS with these short samples of spontaneous production data provides a more meaningful distinction for students in early levels and fails to

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

distinguish well between intermediate and advanced learners. This finding may be expected for the RP, which is intended to be used with elicited data, but it is discouraging for the OGPS, which is intended for use with short samples such as those in this study. To illustrate the point, many students at the institute's Level 5, Level 4, and even Level 3 produce nearly all of the morphosyntactic elements in the OGPS. Yet these students clearly have not reached the highest proficiency possible. Consider the excerpts in (11) and (12) from the speech of C282-3, a Chinese speaker who produces all but two of the morphosyntactic elements in the OGPS.

(11) *It's some different between in the United State and in Taiwan. If I want to buy clothes in the United State, I go to the mall. There are many clothes store in the mall…So it's the different. For food, in the U.S. I only go to grocery store…But in Taiwan I always went to the street to buy food.*

(12) *After a baby is born one month, his parents will give their relatives and close friends some gift. If the baby is a girl, the parents will give them a box of cakes. The custom notice the people, relating people how old the baby is.*

As these excerpts show, some learners at the highest point of development by these measures with these data make a number of errors: syntactic, morphological, and lexical or idiomatic. For instance, prepositions are omitted, agreement endings are incorrect, tense marking is misused, word order in phrases is incorrect, and so on. However, these errors do not factor into these developmental tables because emergence rather than accuracy is used as a measure, and many types of grammatical errors are not included.[6] Additionally, some of the elements that might distinguish between the more advanced learners, such as passives, are rare in the data. It is not clear whether the learners are unable to produce these forms or simply lacked an appropriate context in their response to the prompt.

Toward a New Developmental Measure

There may be a number of ways to give the measures more power to distinguish between intermediate and advanced learners. First, it may be possible to design careful prompts that elicit the desired structures more consistently but are still usable with large groups of learners. Better prompts may improve the power of both the RPS and OGPS analyses. Second, it may be possible to refine the OGPS. This could primarily involve further elaborating on the notion of displacement and simple versus complex subordination. For instance, Young-Scholten et al. (2005) suggested separating subject relative clauses from other relative clauses. They also suggested that it may be possible to use

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

some error data in a limited way, rather than using simple emergence criteria, under the assumption that more advanced learners have fewer errors on the morphosyntactic elements in the developmental measure. For example, rather than simply examining a learner's production data to determine whether she had a single productive use of plural, the percentage of accurate uses of plural could be measured. It is important to note, however, that PT expressly rejects using error data, so this idea would have to be carefully investigated.

Another possibility is to examine variation within a single morphosyntactic element type. In both the OGPS and RPS, the emergence of a single (nonformulaic) element is taken as evidence that the mechanism or structure is available to produce any element of that type. However, not all forms were produced equally in the data. For instance, with respect to modals, *can* appeared earliest, followed by *will* (to indicate future). Learners generally used *could*, *would*, *should*, and *might* later. It is therefore possible that there is an implicational order of modal acquisition within the larger implicational table. This appears to be the case for adverb clauses as well; these learners almost always used the subordinating conjunctions *because* and *if* before *although*, *before*, and *after*. There is some support for the idea of an implicational order within a single structural type; for instance, previous research has indicated that learners may acquire different types of conjunctions as they become more advanced (Kiany, 2001). However, at this point, the cause for these patterns in the data is not clear; it could be a developmental order based on morphosyntactic, phonological, or pragmatic factors or it could merely be reflecting the order of presentation in pedagogical materials. These questions can be addressed through empirical study.

Finally, it is worth noting that certain of the morphosyntactic developments appeared late for most learners: aspect, relative clauses, and passive (from the OGPS) and *wh*-noun clauses (from RP). These elements could serve to provide a better distinction between advanced learners if they were combined in a single measure and elaborated.

Contributing to a New Assessment Measure

As an exploratory step toward creating a practical measure of morphosyntactic development for spontaneous production data, measures were combined from the OGPS and RPS into a single implicational table, which was then maximized for reproducibility and scalability. It was not possible to use all of the measures in the OGPS and RPS to create a table with a C of R over .9 and a C of S over .6; that is, when the morphosyntactic elements from both systems are combined, the table inevitably has a high number of errors. This result may be because these samples

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

of data are not morphosyntactically rich enough or because further refinement is needed for the measures with these data. To create a table with the best predictions for all levels, the measures with the highest error rates in each table were eliminated along with measures that provided no distinction between these learners (i.e., the elements were emerged for every learner). Table 11 displays the results. Note that it is possible that additional morphosyntactic elements could successfully be included with a different group of learners, particularly those at lower levels.

For Table 11, the C of R is .91. This figure indicates a reproducible table. The C of S is .7. A C of S of a minimum of .6–.65 is considered to

Table 11. Implicational table combining measures of RP and OG for maximum scalability

Student	Cop	Co	CC	B/c	Adv 1st	Mod	AC	Do/s	RC	Asp	Pv	Cncl aux
K123-2	*	*	*	N	N	N	N	N	N	N	N	N
A148-2	-*-	*	N	*	N	N	N	N	N	N	N	N
A25-2	*	-*-	-*-	*	N	N	N	N	N	N	N	N
A12-2	N	*	-*-	N	*	N	-*-	N	N	N	N	N
K267-2	-*-	*	-*-	N	*	N	N	**	N	N	N	N
C126-2	-*-	-*-	-*-	-*-	*	N	N	N	N	N	N	N
K269-2	*	*	N	*	*	N	-*-	N	N	-*-	N	N
K266-2	**	-*-	N	-*-	N	-*-	-*-	*	N	N	N	N
K156-2	**	-*-	-*-	N	*	-*-	-*-	N	N	N	N	N
K167-2	**	*	*	*	*	-*-	-*-	N	N	N	N	N
C271-2	**	*	*	-*-	*	-*-	*	N	N	N	N	N
C127-3	**	*	**	*	*	*	N	N	N	N	N	N
K111-3	**	*	-*-	*	*	*	*	N	N	N	N	N
C118-4	**	N	**	*	*	**	*	N	N	N	-*-	N
S366-2	*	-*-	**	-*-	*	*	N	N	*	N	N	N
A45-2	*	*	-*-	*	*	-*-	N	N	-*-	N	N	N
A170-2	**	*	N	-*-	*	**	*	N	-*-	N	N	N
C298-3	*	*	*	N	*	**	*	**	N	N	N	N
K300-3	*	*	-*-	*	*	-*-	*	-*-	N	N	N	N
A338-4	**	*	**	*	*	**	*	**	**	N	N	N
A159-3	*	*	-*-	-*-	*	*	-*-	-*-	N	-*-	N	N
S173-3	**	*	*	*	*	*	**	-*-	N	-*-	N	N
A182-4	*	*	**	N	*	*	**	-*-	*	N	-*-	N
S130-3	**	*	**	*	*	**	-*-	N	-*-	*	N	N
K217-4	**	*	**	N	*	**	**	*	N	N	N	*
K101-3	**	*	*	*	*	**	*	-*-	N	N	N	N
K46-3	**	*	*	-*-	*	N	*	*	N	-*-	N	N
C177-3	**	*	**	*	*	*	*	-*-	N	N	N	N
S362-2	*	*	**	*	*	**	-*-	N	-*-	N	N	N

continued

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Table 11. Continued

Student	Cop	Co	CC	B/c	Adv 1st	Mod	AC	Do/s	RC	Asp	Pv	Cncl aux
A279-4	**	*	**	*	*	**	*	-*-	*	N	N	N
C84-4	*	*	**	*	*	**	**	-*-	**	N	N	N
A157-3	**	*	*	*	*	*	-*-	N	**	N	N	*
C301-3	**	*	**	*	*	**	**	**	*	N	N	N
C282-3	**	*	**	*	*	-*-	-*-	*	*	N	-*-	N
K276-4	**	*	**	*	*	**	**	*	N	-*-	-*-	N
A160-3	*	*	-*-	*	*	*	*	*	-*-	-*-	N	N
T397-5	**	*	**	*	*	*	N	N	*	-*-	-*-	N
A199-4	**	*	**	*	*	-*-	**	-*-	-*-	*	N	N
C201-4	**	*	**	*	*	*	*	-*-	-*-	*	N	N
S391-4	**	*	**	*	*	**	*	*	**	*	N	*
S383-4	**	*	**	*	*	**	-*-	N	*	**	N	*
K320-4	**	*	**	*	*	**	**	**	*	*	N	N
A163-3	**	*	**	*	*	-*-	*	*	-*-	**	N	N
A161-3	**	*	**	*	*	*	**	-*-	-*-	*	N	N
K275-4	*	*	**	-*-	*	*	**	*	-*-	-*-	N	*
S100-5	**	*	**	*	*	**	**	-*-	**	-*-	**	N
J274-5	**	*	**	*	*	**	**	-*-	**	N	-*-	*
C278-4	*	*	**	*	*	**	**	*	*	-*-	-*-	*

Note. Cop = copula forms other than *is*; Co = coordination; CC = complement clauses; B/c = clauses with *because*; Adv 1st = adverb (or prepositional phrase, etc.) as first element in a sentence; Mod = modals; AC = adverb clauses; Do/s = *do* support; RC = relative clauses; Asp = progressive or perfect aspect; Pv = passive; Cncl aux = *wh*-complement clauses.

indicate a scalable table, so it is clear that this is a real progression. Therefore, in these data, morphosyntactic features do emerge in a predictable order for learners of all tested L1s.

However, this order is not supported by either PT or OG and therefore runs counter to the explanatory model of either one. The primary question, then, concerns the basis for this ordering of elements. It seems most likely that a combination of factors is at work, including those in the theoretical proposals: processability (as proposed by Pienemann, 1998, 2005) or the emergence of phrase structure (as proposed by Vainikka & Young-Scholten, 2006). It is also possible that the order is influenced by the complexity of form-meaning matching, order of pedagogical presentation, or frequency. The nature of this combination of factors has not yet been determined.

Other questions remain. For instance, does the pattern of development outlined in this study apply solely to the kind of data collected in RSA-type activities, or can it also be used to describe other types of data (spontaneous or elicited, oral or written)? Does it predict learners' performance on discrete-point tasks (e.g., fill-in-the-blank activities,

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Table 12. Morphosyntactic elements corresponding to ACTFL descriptions

Level	Description	Morphosyntactic elements
Intermediate low	Syntax is strongly influenced by the L1	L1 word order (predicted by OG, not tested in this study)
	Can ask a few questions	Formulaic questions or questions with SVO order (predicted by RP, not tested in this study)
	Can accomplish uncomplicated communicative tasks	Uses subjects, pronouns, copula, coordination
Intermediate mid	Can produce a variety of questions	Do-SV(O)-? (*Do they want food*?) Aux SV(O)-? (*Are they going*?) *Wh*-SV(O)-? (*What they want*?) (predicted by RP, not tested in this study)
	Has difficulty linking ideas	Uses coordination, but not embedded clauses
Intermediate high	Can exchange basic information	*Because* clauses, negation
	Can narrate and describe in major time frames	Modals, especially *will*; past tense emerges for some learners
Advanced low	Can narrate and describe in all major time frames	Modals *will*, *could*, *would* past tense used particularly with copula forms
	Can construct hypotheses, but patterns of error appear	Some adverb clauses, especially those with *if* or *when*; expletive subjects, including *there* and *it*
Superior	Can separate main ideas from supporting information	Adverb clauses, relative clauses (all kinds), noun clauses
	Can construct and develop hypotheses to explore alternative possibilities; abstract elaborations	Clauses with *if, although, until*, etc. Cancel aux-2nd clauses (*wh*-noun clauses); passive; expletive subjects

multiple choice tests)? Can it predict the performance of learners of all L1 backgrounds?

If the order of acquisition presented here for spontaneous production data is corroborated by other research, it may be possible to make important contributions to current assessment measures. This information could primarily serve to supplement the assessment measures that are currently available by providing concrete morphosyntactic benchmarks

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

for learners as they progress. These benchmarks could be used to create an empirically grounded assessment tool that is practical to use with large groups. One way to administer such a test would be to collect small samples of oral data from groups, as was done in this study. It may also be possible to use written data, as in Young-Scholten et al. (2005). The collected responses would then be analyzed for the presence or absence of the features outlined here, and learners could be placed into groups based on their progression through the order of elements.

A second possibility is for these benchmarks to provide objective information to interviewers who are attempting to judge a learner's level by analyzing spontaneous production data. For instance, some of the grammatical terms of the ACTFL scale could be operationalized as in Table 12. This scale would generally not be used with large groups.

Table 12 is an initial attempt to develop a rubric that refers to specific linguistic features and yet is feasible for use as a grammatical assessment tool; however, it would require a good deal of further testing and refinement before it could be implemented.

Further work is clearly needed to refine these proposals, but whatever the outcome, my hope is that this work is a contribution toward a new collaboration, as Read (2007) put it, between SLA and language assessment. My hope is that the collaboration will eventually lead to advances for both fields.

(*Received 10 August 2010*)

NOTES

1. There are some prominent competing hypotheses: for instance, the full transfer/full access hypothesis (Schwartz & Sprouse, 1996), the missing surface inflection hypothesis (Haznedar & Schwartz, 1997; Prévost & White, 2000), and the representational deficit hypothesis (Hawkins & Chan, 1997; Tsimpli, 2003). However, none of these hypotheses outlines a developmental route.

2. This is not to say that there is no L1 influence at all. This theory does allow for L1 influence in identifying cognates, using determiners correctly, and so on. It is also possible that learners with certain L1s could move through the stages more quickly than others. In general, however, effects of L1 influence are not discussed much in Vainikka and Young-Scholten works (1994, 1996a, 1996b, 1998a, 1998b, 2006).

3. All students received the same question, with the exception of two RSAs during the initial launch of the project, when teachers could choose which question to ask. These two RSAs amount to a very small portion of the data (about 6%).

4. Combining several samples of data serves several purposes. First, it simplifies the analysis of data by making comparisons between learners of different levels possible. Second, it ensures that each learner will have a reasonable amount of data for each semester. Third, it mitigates the variations in contexts that occur with differing topic prompts. A more ideal data collection method would perhaps be to have students answer all of the RSA questions in a single sitting so as to take a "snapshot" of each learner's production at a point in time and to mitigate the effects of any progress that learners make over the course of a semester. However, it is unlikely that the time lapse significantly influenced the results; in fact, learners seemed to make very little progress within the course of a semester, based on the measurements in the OGPS and RPS—perhaps an issue to be explored in the future.

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

5. An anonymous *SSLA* reviewer pointed out that it is possible that instruction alters the order of acquisition of structures in the OGPS and RPS, which were originally developed based on data from untutored learners. Note, however, that both sequences have been tested with instructed learners (e.g., Sakai, 2008; Young-Scholten & Ijuin, 2006a, 2006b) and the sequences have been argued to hold.

6. An anonymous *SSLA* reviewer pointed out that advanced L2 learners at the highest stages of the RPS or OGPS may be parallel to children learning English in school from the age of 6 onward; that is, their basic grammar is in place, but they need to learn the sophisticated use of academic language. This is a reasonable possibility, although the parallel breaks down in that advanced L2 learners often are quite inaccurate in their production of many morphological markers, such as tense marking, plurals, and so on, whereas 6-year-olds tend to be quite accurate on these markers.

REFERENCES

ACTFL (American Council on the Teaching of Foreign Languages). (1999). *ACTFL proficiency guidelines: Speaking*. Retrieved April 4, 2010 from http://www.actfl.org/files/public/Guidelinesspeak.pdf

Alhawary, M. (2009). Speech processing prerequisites or L1 transfer? Evidence from English and French learners of L2 Arabic. *Foreign Language Annals*, *42*, 367–390.

Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. (1998). Language testing/SLA interfaces. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 177–195). New York: Cambridge University Press.

Bachman, L., & Cohen, A. (1998). Language testing/SLA interfaces: An update. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 1–31). New York: Cambridge University Press.

Bresnan, J. (2001). *Lexical-functional syntax*. Oxford: Blackwell.

Brindley, G. (1998). Describing language development? Rating scales and SLA. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 112–140). New York: Cambridge University Press.

Chapelle, C. (1998). Construct definition and validity inquiry in SLA research. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32–70). New York: Cambridge University Press.

Clahsen, H. (1990/1991). Constraints on parameter-setting: A grammatical analysis of some acquisition stages in German child language. *Language Acquisition*, *1*, 361–391.

Clahsen, H., & Penke, M. (1992). The acquisition of agreement morphology and its syntactic consequences: New evidence on German child language from the Simone corpus. In J. Meisel (Ed.), *The acquisition of verb placement* (pp. 181–223). Dordrecht: Kluwer.

Cohen, A. (1998). Strategies and processes in test taking and SLA. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 90–111). New York: Cambridge University Press.

Crystal, D., Fletcher, P., & Garman, M. (1976). *The grammatical analysis of language disability*. London: Arnold.

Di Biase, B., & Kawaguchi, S. (2002). Exploring the typological plausibility of processability theory: Language development in Italian second language and Japanese second language. *Second Language Research*, *18*, 274–302.

Douglas, D. (1998). Testing methods in context-based second language research. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 141–155). New York: Cambridge University Press.

Dyson, B. (2009). Processability theory and the role of morphology in English as a second language development: A longitudinal study. *Second Language Research*, *25*, 355–376.

Ellis, R. (2008). Investigating grammatical difficulty in second language learning: Implications for second language acquisition research and language testing. *International Journal of Applied Linguistics*, *18*, 4–22.

Glahn, E., Håkansson, G., Hammarberg, B., Holmen, A., Hvenekilde, A., & Lund, K. (2001). Processability in Scandinavian second language acquisition. *Studies in Second Language Acquisition*, *23*, 389–416.

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Håkansson, G., Pienemann, M., & Sayehli, S. (2002). Transfer and typological proximity in the context of second language processing. *Second Language Research, 18*, 250–273.
Hatch, E., & Farhady, H. (1982). *Research design and statistics for applied linguistics*. Rowley, MA: Newbury House.
Hawkins, R., & Chan, C. (1997). The partial availability of Universal Grammar in second language acquisition: The failed functional features hypothesis. *Second Language Research*, *13*, 187–226.
Haznedar, B., & Schwartz, B. (1997). Are there optional infinitives in child L2 acquisition? In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the annual Boston University conference on language development 21* (pp. 257–268). Somerville, MA: Cascadilla Press.
Jansen, L. (2008). Acquisition of German word order in tutored learners: A cross-sectional study in a wider theoretical context. *Language Learning*, *58*, 185–231.
Kaplan, R., & Bresnan, J. (1982). Lexical-functional grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 173–281). Cambridge, MA: MIT Press.
Kawaguchi, S. (2000). Acquisition of Japanese verbal morphology: Applying processability theory to Japanese. *Studia Linguistica*, *54*, 238–248.
Kiany, G. (2001). On the relationship between English proficiency, writing ability, and the use of conjunctions in Iranian EFL learners' compositions. *ITL Review of Applied Linguistics, 133/134*, 227–241.
Meisel, J., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition*, *3*, 109–135.
Myles, F. (2004). From data to theory: The overrepresentation of linguistic knowledge in SLA. *Transactions of the Philological Society*, *102*, 139–168.
Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam: Benjamins.
Pienemann, M. (2003). *Rapid profile stages*. Retrieved June 22, 2009, from University of Paderborn, Linguistics Department, Rapid Profile Web site: http://groups.uni-paderborn.de/rapidprofile/docs/Stages.pdf
Pienemann, M. (2005). Discussing PT. In M. Pienemann (Ed.), *Cross-linguistic aspects of processablity theory* (pp. 61–83). Amsterdam: Benjamins.
Pienemann, M. (2007a). Processability theory. In B. Van Patten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 137–154). London: Routledge.
Pienemann, M. (2007b, September). *Processability theory: Theory construction and application*. Plenary speech at 17th annual conference of the European Second Language Association, Newcastle, UK.
Pienemann, M., Johnston, M., & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition*, *10*, 217–243.
Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? *Second Language Research*, *16*, 103–133.
Read, J. (2007). Towards a new collaboration: Research in SLA and language testing. *New Zealand Studies in Applied Linguistics*, *13*, 22–35.
Sakai, H. (2008). An analysis of Japanese university students' oral performance in English using processability theory. *System*, *36*, 534–549.
Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, *12*, 40–72.
Shohamy, E. (1994). The role of language tests in the construction and validation of second-language acquisition theories. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second language acquisition* (pp. 133–142). Mahwah, NJ: Erlbaum.
Shohamy, E. (1998). How can language testing and SLA benefit from each other? The case of discourse. In L. Bachman & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 156–176). New York: Cambridge University Press.
Tsimpli, I.-M. (2003). Clitics and determiners in L2 Greek. In J. Liceras, H. Zobl, & H. Goodluck (Eds.), *Proceedings of the 6th generative approaches to second language acquisition conference* (pp. 331–339). Somerville, MA: Cascadilla.
Vainikka, A. (1993/1994). Case in the development of English syntax. *Language Acquisition*, *3*, 257–325.

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press

Vainikka, A., & Young-Scholten, M. (1994). Direct access to X'-theory: Evidence from Korean and Turkish adults learning German. In T. Hoekstra & B. Schwartz (Eds.), *Language acquisition studies in generative grammar* (pp. 265–316). Amsterdam: Benjamins.
Vainikka, A., & Young-Scholten, M. (1996a). The early stages in adult L2 syntax: Additional evidence from Romance speakers. *Second Language Research*, *12*, 140–176.
Vainikka, A., & Young-Scholten, M. (1996b). Gradual development of L2 phrase structure. *Second Language Research*, *12*, 7–39.
Vainikka, A., & Young-Scholten, M. (1998a). The initial state in the L2 acquisition of phrase structure. In S. Flynn, G. Martohardjono, & W. O'Neil (Eds.), *The generative study of second language acquisition* (pp. 17–34). Mahwah, NJ: Erlbaum.
Vainikka, A., & Young-Scholten, M. (1998b). Morphosyntactic triggers in adult SLA. In M. Beck (Ed.), *Morphology and its interfaces in second language knowledge* (pp. 89–113). Amsterdam: Benjamins.
Vainikka, A., & Young-Scholten, M. (2006). The roots of syntax and how they grow: Organic grammar, the basic variety, and processablity theory. In S. Unsworth, T. Parodi, A. Sorace, & M. Young-Scholten (Eds.), *Paths of development in L1 and L2 acquisition* (pp. 77–106). Amsterdam: Benjamins.
Young-Scholten, M., Ijuin, C., & Vainikka, A. (2005, March). *Organic grammar as a measurement of development*. Paper presented at 39th annual TESOL conference, San Antonio, TX.
Young-Scholten, M., & Ijuin, C. (2006a, March). *Expanding organic grammar-based assessment*. Paper presented at 40th annual TESOL conference, Tampa, FL.
Young-Scholten, M., & Ijuin, C. (2006b, September). How can we best measure adult ESL student progress? *TESOL Adult Education Interest Section Newsletter*, *4*(2).

APPENDIX

RSA Questions

(1) *How do you feel about pets? Do many people have pets in your country? How are they treated, in general?*
(2) *Talk about a person who was very important to you in the past. Who was this person? Why was this person important to you?*
(3) *What is the biggest problem your country is facing today? How would you change it?*
(4) *Is shopping for food in your country the same as in the US? Explain how it is different and how it is the same.*
(5) *Describe something that you liked to do when you were in your country but that you can't do here. Where did you do this? Why did you like it? How did it make you feel?*
(6) *Choose a custom (baby's birth, wedding, funeral, entry to adulthood, etc.) in your country. Describe what is done for this custom and why.*
(7) *Talk about a famous person, place, or event in your country.*
(8) *What sports do you enjoy? What do you like about them?*
(9) *Talk about an important event that happened in the past in your country.*
(10) *Talk about a place that you really like. Describe it and tell why you like it.*
(11) *Talk about a funny or scary experience that you had.*
(12) *Talk about your favorite holiday. What do you like about it?*

https://doi.org/10.1017/S0272263111000301 Published online by Cambridge University Press